

A novel feature selection method to boost variable predictive model-based class discrimination performance and its application to intelligent multi-fault diagnosis

Measurement and Control

1–15

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0020294019877497

journals.sagepub.com/home/macSongrong Luo^{1,2,3} , Wenxian Yang³ and Hongbin Tang^{3,4}

Abstract

Effective and efficient incipient fault diagnosis is vital to the maintenance and safe application of large-scale key mechanical system. Variable predictive model-based class discrimination is a recently developed multiclass discrimination method and has been proved to be potential tool for multi-fault detection. However, the vibration signals from dynamic mechanical system always present non-normal distribution so that the original variable predictive model-based class discrimination might produce the inaccurate outcomes. An improved variable predictive model-based class discrimination method is introduced at first in this work. At the same time, variable predictive model-based class discrimination will suffer computation difficulty in the case of high-dimension input features. Therefore, a novel feature selection method based on similarity-fuzzy entropy is presented to boost the performance of the variable predictive model-based class discrimination classifier. In this method, the ideal feature vectors are optimized to acquire more accurate similarity-fuzzy entropies for the input features. And, the one with the largest similarity-fuzzy entropy value is removed to refine input feature subset. Moreover, the optimal input features are repeatedly evaluated using the improved variable predictive model-based class discrimination classifier until the expected results are achieved. Finally, the incipient multi-fault diagnosis model for a hydraulic piston pump is established and verified by experimental test. Some comparisons with commonly used methods were made, and the results indicate that the proposed method is more effective and efficient.

Keywords

Similarity-fuzzy entropy, feature selection, variable predictive model-based class discrimination, intelligent multi-fault diagnosis, hydraulic pump

Date received: 18 June 2019; accepted: 20 August 2019

Introduction

The sudden failure of key engineering equipment, such as wind turbine and construction machine, will likely lead to unexpected break down and even casualties resulting in huge financial loss. Therefore, it will be plausible if any fault in them can be detected automatically as soon as possible. However, the reliable and real-time intelligent condition monitor and fault diagnosis is still a challenge due to the complex structure and work mechanism with harsh working environment and heavy working load for these key engineering equipments.^{1,2}

Generally speaking, intelligent fault diagnosis is essentially pattern recognition³ which includes feature extraction, feature selection, and pattern recognition.⁴ So far, many data-driven intelligent fault diagnosis techniques have been developed.^{5–12} Many powerful

signal processing techniques, such as time-domain statistic analysis, wavelet transform,¹³ and empirical mode decomposition (EMD) and its extension methods,^{14,15} have been employed to extract sensitive fault features from different domains. Subsequently, to achieve accurate results, a feature selection technique is used to

¹Cooperative Innovation Center for the Construction & Development of Dongting Lake Ecological Economic Zone, Changde, China

²College of Mechanical Engineering, Hunan University of Arts and Science, Changde, China

³School of Engineering, Newcastle University, Newcastle upon Tyne, UK

⁴College of Automotive and Mechanical Engineering, Changsha University of Science & Technology, Changsha, China

Corresponding author:

Songrong Luo, School of Engineering, Newcastle University, Newcastle upon Tyne NE1 1RU, UK.

Email: luosongrong@126.com



reduce the dimension of features before applying various pattern recognition techniques to identify different situations.^{16–23} Manifold learning has been combined with Shannon wavelet support vector machine (SW-SVM) to fulfill the fault diagnosis for a wind turbine transmission system.²⁴ The hierarchical symbol dynamic entropy was extracted to quantify the complexity of signals caused by early faults of rolling bearings, and the binary tree support vector machine (BT-SVM) served as intelligent recognition approach.²⁵ But, the widely used pattern recognition methods, such as artificial neural network (ANN) and support vector machine (SVM), have their individual shortcomings. For example, the ANN method needs plenty of samples and has low speed due to complex iterating computation; SVM is a binary classifier and requires rigid parameter tuning so that it is complicate to solve the multiclass problem. Variable predictive model-based class discrimination (VPMCD) is a novel multiclass discrimination method.^{26–28} The VPMCD can make full use of interactions among the features to establish mathematical models—variable predictive models (VPMs)—for each class so as to identify the classes without complex iterating nor strict parameters adjustment.^{3,29–31} Recently, many works have shown that the VPMCD classifiers have much better performance and are a potential tool for multiclass fault diagnosis.^{29,32,33} However, the original VPMCD method cannot always adapt the non-normal distribution characteristics of features collected from vibration signals in dynamic mechanical system. Hence, the improved VPMCD technique is introduced in this paper. The details can be found in section “VPMCD principle and improvement strategy.”

Moreover, it is found that the computation expense of the improved VPMCD classifiers will greatly increase when the feature dimension is larger to acquire the richer information, which would limit its efficiency for multiclass discrimination. If some irrelevant features are removed using a feature selection technique, then the information which is sensitive to fault type and fault degree for effective classification is presented with fewer features, then the improved VPMCD classifier would have less time cost and higher accuracy. In general, feature selection algorithms include filter models, wrapper models, and embedded models.^{19,20,22,23,34,35} Filter models evaluate the general characteristics of the training data to select a feature subset without employing any learning algorithm; thus, it has less computation cost.²³ Nevertheless, it might obtain the feature subsets irrelevant to classes. Recently, Tang et al.¹⁹ proposed a feature selection method and applied it to bearing fault identification to improve the VPMCD performance, but the procedure is very complicated and difficult to be applied in practice. The wrapper models assess the performance of the feature subset selected using a pre-determined classification algorithm so that the classifiers can usually achieve more accuracy, but they are usually slower than filter models.²³

Fuzzy entropy based on similarity measure can indicate the relevance of features and classes and has been used to successfully complete the wrapper feature selection with similarity classifier.^{22,35} However, the similarity measure was usually computed by the mean value, likely leading to inaccurate results. Simultaneously, the similarity classifier is complicated for the parameters adjustment. A novel wrapper feature selection technique integrating similarity-fuzzy entropy with the improved VPMCD was proposed to refine input features so as to boost the identification efficiency in the intelligent multi-fault diagnosis in this work. And, the proposed method was applied to establish the incipient intelligent multi-fault diagnosis model including three types of single-fault and two types of multi-fault for a hydraulic piston pump in a construction machine. The experimental results show that the intelligent multi-fault model with the novel feature selection technique and the improved VPMCD is more effective and efficient.

The rest of the paper is written as follows. The VPMCD method and its improved technique are introduced in section “VPMCD principle and improvement strategy.” A novel wrapper feature selection technique which integrated the improved similarity-fuzzy entropy and the improved VPMCD is developed in section “A novel feature selection method.” Considering that piston pumps play great role in hydraulic system of some key large-scale equipments, the proposed technique is applied to incipient intelligent multi-fault diagnosis and verified in section “Application to incipient intelligent multi-fault diagnosis for piston pumps” and the conclusions are drawn in section “Conclusion.”

VPMCD principle and improvement strategy

Basic VPMCD method

The feature variables extracted from the original signals and their intrinsic relationship can always quantify the natural characteristics of the dynamic system. Based on this hypothesis, a new multiclass discrimination method, called variable prediction model class discrimination (VPMCD), has been proposed and applied to medicine signal analysis.^{26–28} The VPMCD can establish mathematical variable prediction models (VPMs) to discover the intrinsic and quantitative interactions among these feature variables and utilize these VPMs to identify the classes of unknown test samples. The VPMCD is implemented in the following two steps described as Figure 1.

Step 1: VPM training. Suppose we collected g groups training samples extracted p different feature variables, expressed as a vector $X = [x_1, x_2, \dots, x_p]$ for each sample. Given there are classes together and n training samples are collected, so training sample matrix $TS [n \times p; g]$ can be obtained. This training matrix TS is divided into g submatrices $G_k [n_k \times p]$, each of which

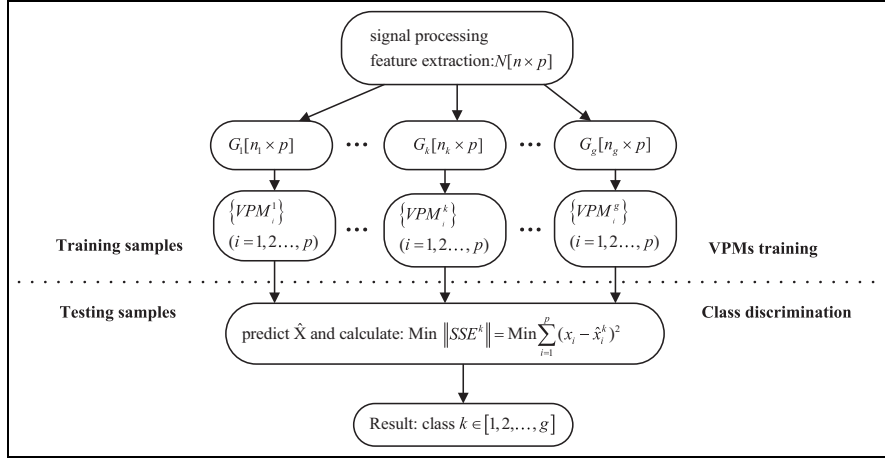


Figure 1. Flowchart of the VPMCD algorithm.

Table 1. Details on design matrix and number of parameters.

Model type	Design vector, D_i^k	Number of parameters, q
L	$[1, x_1, x_2, \dots, x_r]$	$1 + r$
LI	$[1, x_1, x_2, \dots, x_r, x_1 x_2, x_1 x_3, \dots, x_{r-1} x_r]$	$1 + r + C_r^2$
QI	$[1, x_1, x_2, \dots, x_r, x_1^2, x_2^2, \dots, x_r^2, x_1 x_2, x_1 x_3, \dots, x_{r-1} x_r]$	$1 + 2r + C_r^2$
Q	$[1, x_1, x_2, \dots, x_r, x_1^2, x_2^2, \dots, x_r^2]$	$1 + 2r$

L: linear; LI: linear + interaction; QI: quadratic + interaction; Q: quadratic.

served as a special class k , where n_k is the number of the training samples for class k ($k = 1, 2, \dots, g$), that is, $\sum n_k = n$. For class k , the feature variable vector is noted $X_k = [x_1^k, x_2^k, \dots, x_p^k]$. The mathematical predictive model VPM_i^k for any feature variable x_i^k can be established with either a linear or a nonlinear regression equation in the form of formulas (1)–(4).¹⁷ In other words, the function VPM_i^k can identify the interactions between the variable x_i^k and the other variables x_j^k ($j = 1, 2, \dots, r, j \neq i$) in the same class k , and the set of other feature variables is termed as the predictor variable set. The element number of the predictor variable set is referred as predictor order and their different combination form is called as the design vector D_i^k , which is determined by the regression model type and the predictor order r . The regression models used mostly are listed below.

Linear (L) VPM

$$X_i = b_0 + \sum_{j=1}^r b_j X_j \quad (1)$$

Linear + Interaction (LI) VPM

$$X_i = b_0 + \sum_{j=1}^r b_j X_j + \sum_{j=1}^r \sum_{k=j+1}^r b_{jk} X_j X_k \quad (2)$$

Pure Quadratic (PQ) VPM

$$X_i = b_0 + \sum_{j=1}^r b_j X_j + \sum_{j=1}^r b_{jj} X_j^2 \quad (3)$$

Quadratic + Interaction (QI) VPM

$$X_i = b_0 + \sum_{j=1}^r b_j X_j + \sum_{j=1}^r b_{jj} X_j^2 + \sum_{j=1}^r \sum_{k=j+1}^r b_{jk} X_j X_k \quad (4)$$

Supposed there are g classes and p feature variables in each class, then the model matrix VPM including $g \times p$ elements can be established using feature vectors $TS[n \times p; g]$ of the training samples in the first step. Noted that for any feature x_i^k , there are g VPMs, written as $\{VPM_i^k (k = 1, 2, \dots, g)\}$. Once the model type and the order r are selected, $\{VPM_i^k\}$ can be obtained using the known submatrices $G_k[n_k \times p]$ from the training samples.

In particular, the VPM_i^k can be obtained by solving the matrix equation $D_i^k B_i^k = x_i^k$. Herein, the design vector D_i^k is listed in Table 1. B_i^k is a coefficient vector which can be obtained by solving the regression equation with the aid of the least squares regression method. The least square method finds the regression coefficient by taking the minimum of the squared error of the fitting as the objective function.

Once the coefficients are estimated, the set of best predictive models VPM_i^k belonging to class k will uniquely characterize the variable associations for the class k . The individual class models VPM_i^k are represented by model coefficient vector B_i^k and set of predictive variables formulating the design vector D_i^k .²⁶⁻²⁸ A structure of VPM can be symbolically shown using matrix representations (5), each array vector of which represents the VPMs for each class k .

It is worthwhile pointing out that after choosing the model type and its order r , the number of the specific forms of VPM_i^k is $d = (p - 1)C_r$. For example, when LI model and $r = 2$ are selected and 20 features are extracted from the training samples, then $p = 20$, and there are $d = (20 - 1)C_2 = 171$ possible VPM_i^k models. Whereas, only the best VPM_i^k model would be selected out of the d possible models by the minimum sum of squared prediction errors. The best model is the one with minimum sum of squared prediction errors

$$VPM = \begin{bmatrix} VPM_1^1 & VPM_2^1 & \cdots & VPM_p^1 \\ VPM_1^2 & VPM_2^2 & \cdots & VPM_p^2 \\ \vdots & \vdots & \ddots & \vdots \\ VPM_1^g & VPM_2^g & \cdots & VPM_p^g \end{bmatrix} = \begin{bmatrix} D_1^1 B_1^1 & D_2^1 B_2^1 & \cdots & D_p^1 B_p^1 \\ D_1^2 B_1^2 & D_2^2 B_2^2 & \cdots & D_p^2 B_p^2 \\ \vdots & \vdots & \ddots & \vdots \\ D_1^g B_1^g & D_2^g B_2^g & \cdots & D_p^g B_p^g \end{bmatrix} \quad (5)$$

Step 2: class discrimination. As dedicated above, each VPM_i^k stores the design vector and the coefficient vector. For an unknown sample, using feature extraction, the new feature variable vector $X_{test}[1 \times p]$ which is written as $[x_{t1}, x_{t2}, \dots, x_{tp}]$ is obtained, then these trained models VPM_i^k are used to calculate the predictive values \hat{x}_{ti}^k of any features variable x_{ti} , respectively. The predictive value matrix \hat{X}_{test} of feature variable can be accomplished by projecting the new feature variable vector $X_{test}[1 \times p]$ on these VPM_i^k . The predictive value matrix \hat{X}_{test} is expressed as formula (6)

$$\hat{X}_{test} = \begin{bmatrix} \hat{x}_{t1}^1 & \hat{x}_{t2}^1 & \cdots & \hat{x}_{tp}^1 \\ \hat{x}_{t1}^2 & \hat{x}_{t2}^2 & \cdots & \hat{x}_{tp}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_{t1}^g & \hat{x}_{t2}^g & \cdots & \hat{x}_{tp}^g \end{bmatrix} \quad (6)$$

Moreover, the sum SSE^k of the prediction errors can be easily obtained as

$$SSE^k = \sum_{i=1}^p (x_i - \hat{x}_i^k)^2 \quad (7)$$

Therefore, the error square sum vector $[SSE^1, SSE^2, \dots, SSE^k, \dots, SSE^g]$ would be definitely completed. If SSE^k is the minimum one element of the error square sum vector, then the unknown sample

certainly comes from class k and thus the class discrimination is achieved. VPMCD classifier can discriminate the class of an unknown test sample according to the discrimination function given in equation (8), and the test sample is classified to class k when minimum value SSE^k has been met. The further details can refer to the literature Raghuraj and Lakshminarayanan²⁷

$$\text{Min}SSE^k = \sum_{i=1}^p (x_i - \hat{x}_i^k)^2; k = 1, 2, \dots, g \quad (8)$$

In summary, the multivariate interactions expressed by the VPMs can be mathematically established, and these mathematical models will show distinct dissimilarities among the classes. Moreover, the distinct relations in the VPMs are specific to each class so that they can be directly used as class discrimination models. As a model-based multiclass discrimination approach, the VPMCD technique has advantages of robustness with less computational effort and is greatly useful to multiclass problem.

Improved VPMCD

As mentioned above, the coefficient vector B_i^k of VPM can be computed by the least square algorithm for the regression equation. However, it is not always able to achieve stable and reliable outputs when it is applied to the vibration signals. First, the vibration signals are not always normal-distributed. But the basic VPMCD ignores the possibility. As a result, the established VPMs are heteroscedastic, leading to reduction in the recognition rate greatly. In addition, the model structure for VPM mainly depends on the nature of the dynamic system and the variables selected. However, the basic VPMCD determines the signal model structure by selecting artificially, which would result in unstable outputs. Targeting these issues above, two aspects of improvements are made in this paper.

One aspect of improvement is that the weighted least square (WLS) algorithm is utilized to compute the coefficient vector of VPMs instead of the normal least square technique. The training process for VPMs in the VPMCD method is actually a linear or nonlinear regression procedure, and the estimation method of the model parameters is extremely significant for the accuracy and stability of the VPMCD method. In the basic VPMCD method, the least square algorithm is used to solve the issue. As we all know, the premise of the least squares method is that the variables are independent and subject to a normal distribution. However, in the practice of mechanical fault diagnosis, the error of measurement data tends to be non-normal distribution. With least squares estimation, the regular equation will be ill-conditioned, and the effective parameter estimation value may not be obtained. Moreover, when using the least squares method, in order to improve the stability of the parameter estimation, the sample capacity is often required to be high, and the general requirement

is greater than 30 or greater than five to eight times the number of variables. But, it is difficult to meet the requirement since fault samples are often not easily acquired resulting in rare fault samples in mechanical fault diagnosis. Therefore, the WLS algorithm is used to obtain more reliable and accurate parameters for VPMs in the VPMCD method.

Unlike the general least square algorithm, the WLS algorithm computes the regression coefficient by taking the minimum of the weighted squared error of the fitting as the objective function, which is expressed as

$$\sum_{i=1}^n w_i \left| Y_i - \sum_{j=1}^p x_{ij} \beta_j \right|^2 = \min \quad (9)$$

where w_i is the weight and β_j is the values of the regression coefficient.

The other aspect of improvement is an ensemble VPMCD strategy. Since the basic VPMCD has to decide the model type and its order r before VPMs training and use single model type to establish the VPM, it easily causes unstable classification. Here, we employed an ensemble technique to target this issue.

First, for a feature variable x_i^k in a certain class k , N models VPM_{in}^k ($n = 1, 2, \dots, N$) can be obtained by the combination of different model types (L, LI, Q, or QI) and different model orders r to reflect different interactions among the feature variables from different viewpoints. Second, N class discrimination results can, respectively, be acquired using these classifiers designed by VPM_{in}^k . At last, the vote technique is used to decide the final results. In general, the number N of the models is suggested to set as 3–5 to balance the computation tense and classification rate. In this paper, four types of model were used to get VPM_{in}^k individually. They are L model with $r = 1$, LI model with $r = 2$, Q model with $r = 1$, and QI model with $r = 2$; the first two models mainly capture the linear relationship, while the latter two are basically for nonlinear interactions.

In summary, the VPMCD would be flexible to the data distribution of features extracted from the nonlinear and non-stationary vibration signals of mechanical dynamic system after improvement. However, the number of the possible VPMs equals to $d = (p - 1)C_r$ in the training procedure. That means, when we extract lots of features to describe the dynamic characteristics from different domains, the calculation time cost would greatly increase. This would diminish the advantages of the VPMCD multiclass classifier. Actually, many feature variables are less irrelevant or they are less important to the classes. In practice, the computation is quite slow when the feature dimension becomes more than eight. If the irrelevant feature variables are picked out and only more effective features are reserved, the classifiers' speed will be faster. Therefore, we propose a novel feature selection method to refine more relevant features for VPMCD classifiers. The proposed method and its details are described as section "A novel feature selection method."

A novel feature selection method

Fuzzy entropy brief

The basic principle in similarity classifier is calculating the similarity measure between the observations and an ideal vector V and then deciding which class the sample belongs to according to the similarity measure. If we get the similarity value between the sample $x = [x_1, x_2, \dots, x_m]$ and the ideal vector V of class i being 1, the sample belongs to class i . If we get the similarity value being zero, the sample does not belong to class i . Now, this idea is introduced to check feature relevance using fuzzy entropy based on similarity. The small entropy values represent formativeness, whereas the high entropy values show signals uncertainty. m similarity measures are calculated (where m is the number of features); if the similarity is 1 or 0, then the formation is certain and the fuzzy entropy is low; if the similarity is close 0.5, then the formation is high uncertain and the fuzzy entropy is high, vice versa. Therefore, the fuzzy entropy measures are good tools to indicate the relevance of features in a data set and have been used for feature selection.^{21,22,35,36} When the similarities between ideal vector value of this feature and sample vector values of this feature are calculated, we can find the feature with the highest entropy, which is the lowest relevant feature to the classes and should be removed to reduce the feature dimension.²⁵ In our work, we use the newer measure of fuzzy entropy to quantify the relevance, which is defined as^{22,35,36}

$$H(A) = \sum_{i=1}^n \left[\sin \frac{\pi \mu_A(x_i)}{2} + \sin \frac{\pi(1 - \mu_A(x_i))}{2} - 1 \right] \quad (10)$$

where $\mu_A(x_i)$ is the fuzzy set. $\mu_A(x_i)$ refers to the similarity measure of each feature in the proposed method.

Feature selection based on similarity-fuzzy entropy with the improved VPMCD

Similarity measure indicates the similar degree between the sample X and the ideal vector V in class i . After obtaining the training sample set X , the similarity S of feature d of an observation x_j of the training sample set with that of the idea vector V in class i is expressed as^{22,35}

$$S(x_{j,d}, v_{i,d}) = 1 - |x_{j,d} - v_{i,d}| \quad (11)$$

The similarity can be acquired for each training sample x_j in the training sample set X for each feature d and class i . If similarity is 1 or 0, the fuzzy entropy value is low, showing high in formativeness; whereas if similarity is close to 0.5, its fuzzy entropy is the highest, indicating signal uncertainty. In other words, if fuzzy entropy is higher, then the feature is less relevant for the classification procedure. Therefore, in order to reduce the amount of irrelevant features, similarity

measure and fuzzy entropy were combined to develop a feature selection algorithm for medical data sets.^{22,35} But in these literature works, the similarity measures calculated by the mean values of features served as the input vector to calculate the fuzzy entropy value, and classification results were acquired using similarity classifier. This would bring two shortcomings. First, the similarity classifier is actually based on distance measures, and its parameters have significant influence on clustering results,^{22,35} while it is complicated to adjust them. In addition, when calculating similarity measures of features, the generalized mean values of features were used for ideal feature vectors, which would lead the ideal vector to change with different training samples used. In other words, when training samples used are different, the removal features will vary much, likely leading to unreliable results for feature selection. Therefore, in our proposed feature selection technique, fuzzy C-mean (FCM) cluster is utilized to provide the more accurate ideal feature vectors. As we all know, FCM is a soft clustering approach and can meet objective function using iterating optimization, so the stable cluster centers can be obtained as the ideal feature vector. Meanwhile, the improved VPMCD method served as a multiclass discrimination approach. The flowchart of proposed method is shown in Figure 2, and the details are illustrated as follows:

1. Divide the collection samples set into the training sample subset and the test sample subset randomly, and use the FCM to compute the cluster centers c_i for the i th class. The related expressions used in the FCM can be found in the following

$$c_i = \frac{\sum_{j=1}^n \mu_{ij}^g x_{ij}}{\sum_{j=1}^n \mu_{ij}^g} \quad (12)$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{1}{g-1}}} \quad (13)$$

$$d_{ij} = |v_j - x_i| \quad (14)$$

FCM optimizes the objective function to obtain the optimal cluster centers. Its objective function is provided by

$$J = \sum_{i=1}^C J_i = \sum_{i=1}^C \sum_{j=1}^N \mu_{ij}^m d_{ij}^2 \quad (15)$$

where μ_{ij} is the degree of membership for the i th sample in j th class, $g \in [1, \infty)$ is the weight coefficient. C is the number of classes and n is the number of training samples. The optimal cluster centers represent the average character of each class, which can be regarded as the best representative point of this class. So, the optimal cluster centers will be regarded as the ideal vectors

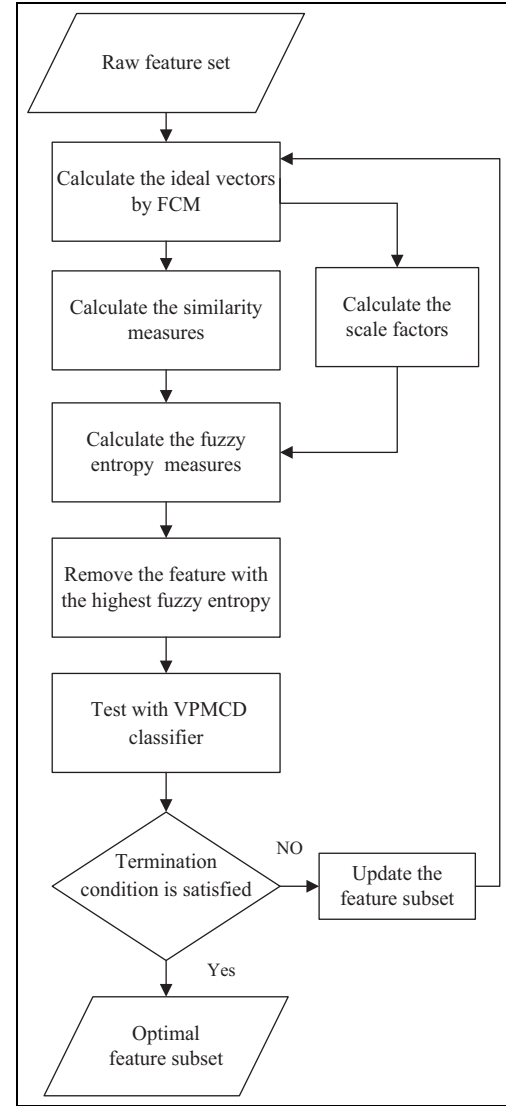


Figure 2. Flowchart of the proposed feature selection method.

and ideal vectors $V = \{v_1, v_2, \dots, v_C\}$ with C elements can be determined in the end.

2. Calculate the similarity S with scale factor for each feature of training samples in each class. Supposed there are n training samples, p features, and C classes, the result with $C^*(mn)$ matrix will be output. In the procedure, to emphasize the distance between the ideal vectors of the classes, the scale factor is introduced to compute the modified fuzzy entropy for each feature in each class. The scale factor is defined as²²

$$S_{i,d} = 1 - \frac{\sum_{i \neq j} |v_{i,d} - v_{j,d}|}{C-1} \quad (16)$$

Then, the similarity measures with scale factor can be calculated as

$$S = S(x_{j,d}, v_{i,d}) \times S_{i,d} \quad (17)$$

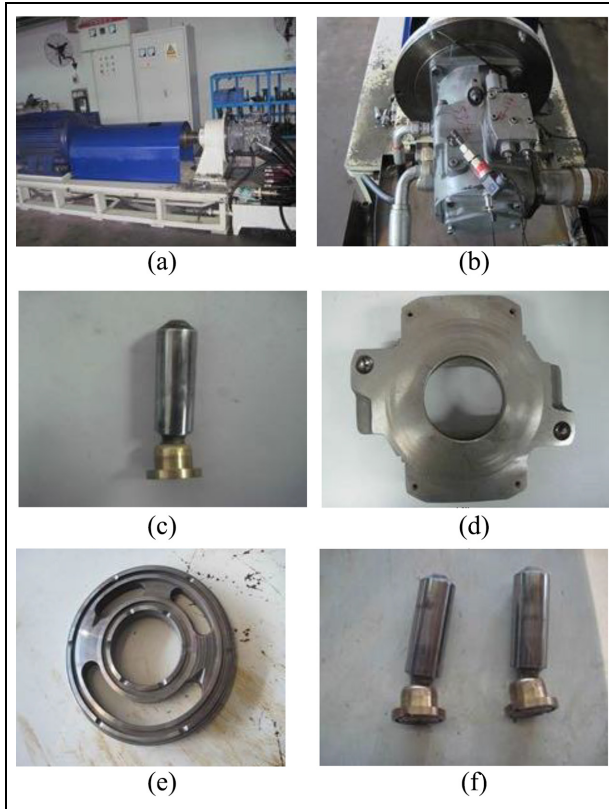


Figure 3. Experiment rig of hydraulic pump and the major parts of the axial piston pump: (a) experiment test rig, (b) axial piston pump, (c) single piston, (d) swash plate, (e) valve plate, and (f) counter-position pistons.

3. Calculate the fuzzy entropy measures and remove the less relevant features to fulfill feature selection. The fuzzy entropy can be acquired using the similarity measures S as the fuzzy set value by formula (10). Then, the feature with the highest entropy will be removed, and the new feature subset will be tested with the VPMCD classifier, and all steps are repeated the feature subset is updated until satisfactory performance is achieved and the optimal feature subset is completed.

Application to incipient intelligent multi-fault diagnosis for piston pumps

Piston pumps play important role for a hydraulic system in aircraft, construction machine, and so on. The major parts of a piston pump include in cylinder block, driving shaft, piston, swash plate, valve plate, and slipper and have three most crucial frictional pairs: piston and cylinder bore, swash plate and slipper, and valve plate and cylinder block. When a piston pump is running, due to load fluctuation and impact excitation, the frictional pairs are easy to wear out, which will lead to incipient abrasion failures. Therefore, efficient real-time incipient fault diagnosis for piston pumps is essential to guarantee high reliability and safety.^{16,37–40} In this

section, an incipient fault diagnosis model for piston pumps was developed and applied using the proposed feature selection method with the improved VPMCD.

Data set collection

The experimental vibration signals were collected from axial piston pump experiment rig. The experimental rig and the relevant details are shown in Figure 3. The axial piston pump used is A11VLO190 with nine pistons. The parts were obtained from the fault part library. The speed of driving shaft was 1600 r/min, and the pressure of the main hydraulic circuit was kept at 10 MPa. The accelerometer was installed to the axial piston pump with magnetic base to collect the vibration signals using NI9233 data acquisition card with a sampling frequency of 10 kHz. The piston pump was tested under six different conditions. They are noted as class 1–class 6: class 1—normal condition, class 2—one piston abrasion, class 3—swash plate abrasion, class 4—valve plate abrasion, class 5—both swash plate and valve plate abrasion, and class 6—counter-position pistons abrasion. There were obviously three types of single-fault and two types of multi-fault. The vibration signals under different conditions and their fast Fourier transformation (FFT) spectrum are given in Figures 4 and 5, respectively. We collected 60 raw vibration signals under each condition and 360 raw vibration signals in total. The time of each sample was 0.25 s.

Feature extraction

From Figures 4 and 5, it was found that the vibration signals collected under different conditions had no distinct difference in structure from each other. But, compared with normal condition, the vibration intensity of the signals under various abnormal conditions increases to different degrees; meanwhile, both the distribution of frequency components and energy change greatly. In other words, the signals show different complexities in structure. In order to realize intelligent fault diagnosis, the raw data sets were processed and some features were extracted for characterizing the signals from different domains. Here, 15 features were obtained using some useful techniques such as time-domain statistic, singular value decomposition (SVD),⁴¹ and local characteristic-scale decomposition (LCD).⁴² The first 10 ones were widely used time-domain statistic indexes, they are listed as follows: (1) peak, (2) mean square value, (3) variance, (4) standard deviation, (5) root mean square (RMS), (6) waveform metric, (7) peak metric, (8) pulse metric, (9) skewness, and (10) kurtosis. Then, the special five features from different viewpoints were extracted shown as follows.

From Figure 5, it was observed that when the running condition of the piston pump was normal, the frequency region of the vibration energy ranged from 1900 to 2200 Hz, and when faults happened, the vibration energy moved toward the lower frequency region.

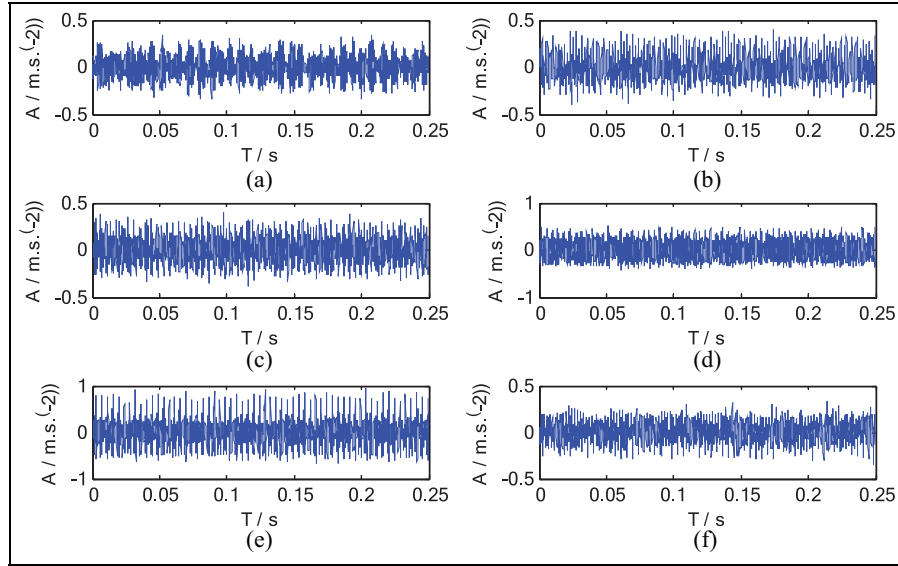


Figure 4. Time-domain waveform: (a) normal condition, (b) one piston abrasion, (c) swash plate abrasion, (d) valve plate abrasion, (e) both swash plate and valve plate abrasion, and (f) counter-position pistons abrasion.

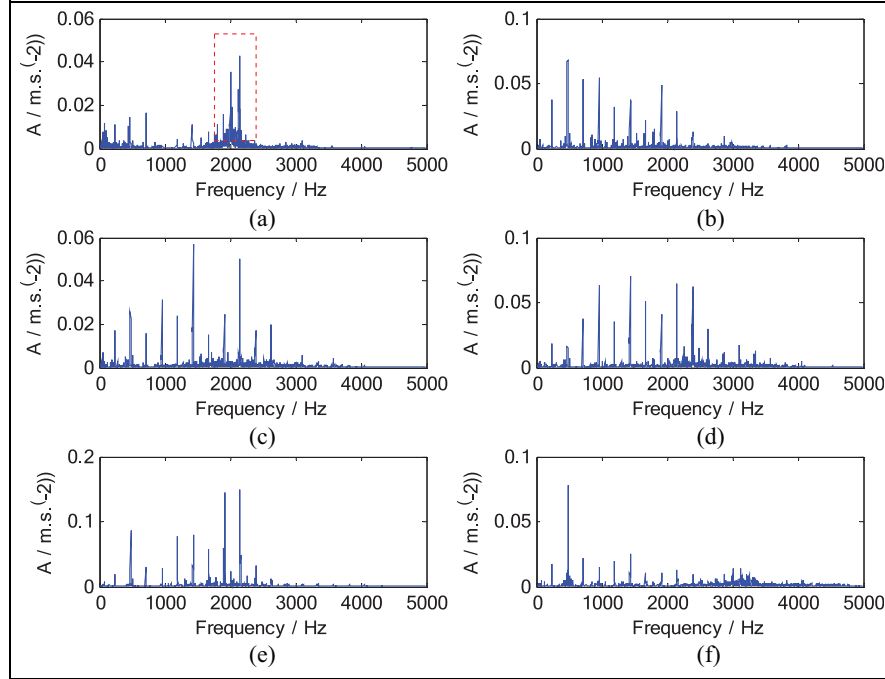


Figure 5. FFT spectrum: (a) normal condition, (b) one piston abrasion, (c) swash plate abrasion, (d) valve plate abrasion, (e) both swash plate and valve plate abrasion, and (f) counter-position pistons abrasion.

So, the 11th and the 12th features to indicate the change for energy distribution with running conditions were designed as follows:

11. The ratio of the vibration in the frequency region⁴³ (1900 Hz, 2200 Hz)

$$R_{f1} = \frac{\sum_{i=a}^b d_i}{\sum_{j=1}^N d_j} \quad (18)$$

where d_i ($i = 1, 2, \dots, N$) are the spectrum value, and a and b are the number of the first and the last points in spectrum data in the frequency region (1900 Hz, 2200 Hz), respectively.

12. The ratio of the vibration in the frequency region⁴³ (1 Hz, 1900 Hz)

$$R_{f2} = \frac{\sum_{i=1}^a d_i}{\sum_{j=1}^N d_j} \quad (19)$$

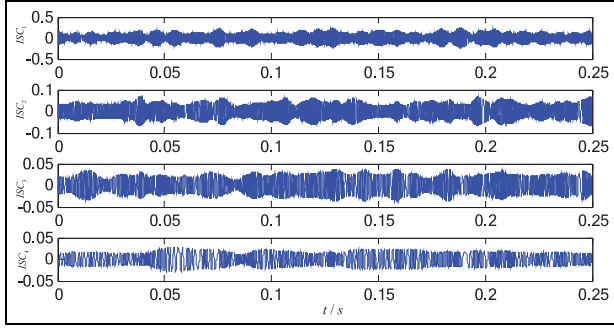


Figure 6. LCD results of the signal under normal condition.

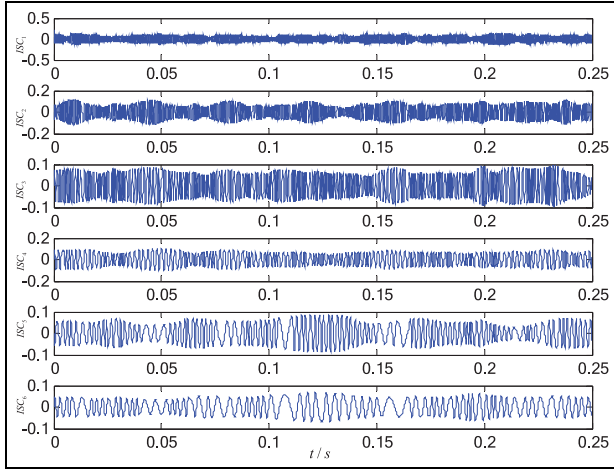


Figure 7. LCD results of the signal under one piston abrasion.

13. Singular spectrum entropy, E_{svd} .

SVD using matrix reconstruction is an important analytical method for time series.^{41,44} However, the reconstruction parameters greatly influence the result of SVD method, so that it is difficult to determine them.⁴⁴ Hence, a special skill was used to deal with the issue as follows. First, we used LCD to transform the collected signal into the set of intrinsic scale components (ISCs)⁴²

$$x(t) = \sum_{i=1}^N ISC_i(t) + r(t) \quad (20)$$

where $ISC_i(t)$ is the i th component and $r(t)$ is the residual component. Second, the matrix $\{ISC_i(t), i = 1, 2, \dots, m\}$ was analyzed by SVD method and m singular values $\{\lambda_i, i = 1, 2, \dots, m\}$ can be obtained, which depicted the contribution of individual ISC to the complexity and uncertainty of the original signal.⁴² In this experiment, the normal signal was decomposed into four ISCs shown in Figure 6, while the signals under different fault conditions were decomposed into six to nine ISCs. The first six ISCs under different conditions are shown in Figures 7–11. From Figures 6–11, it was found that the frequency bands and energy distribution of ISCs were varying under various running conditions.

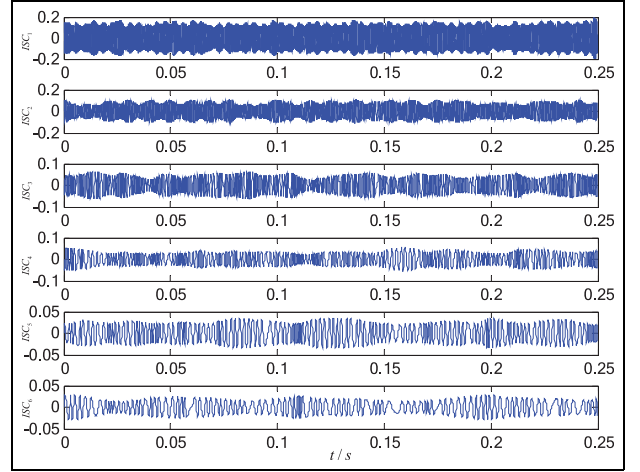


Figure 8. LCD results of the signal under swash plate abrasion.

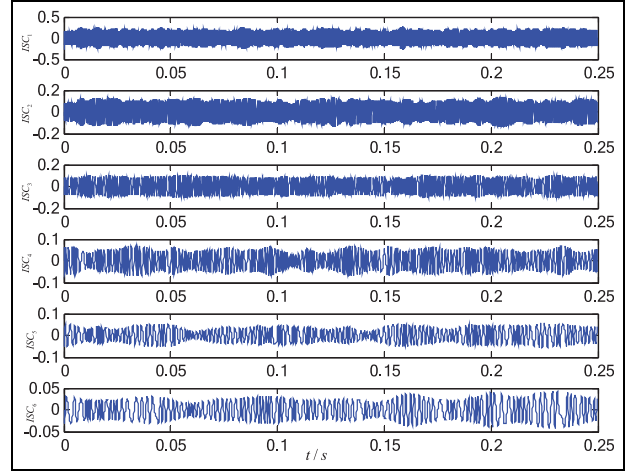


Figure 9. LCD results of the signal under valve plate abrasion.

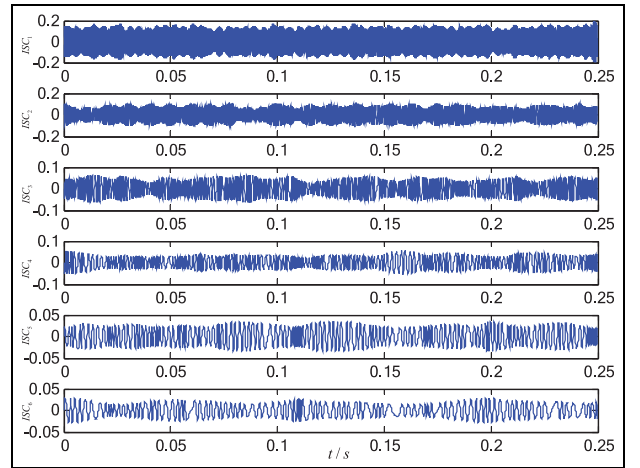
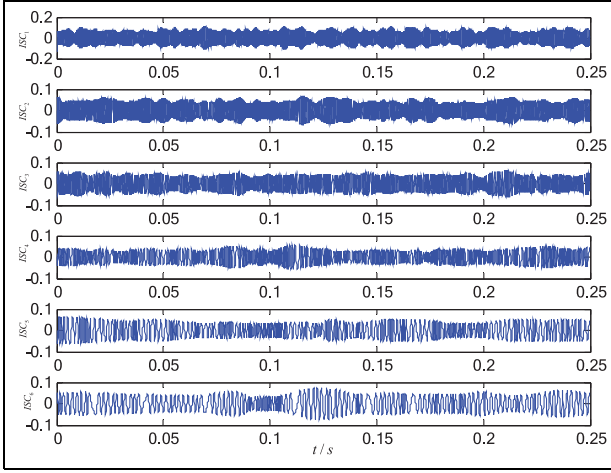


Figure 10. LCD results of the signal under both swash plate and valve plate abrasion.

Recently, related researches have indicated that entropy-based measures are powerful information extraction tool for analyzing complex time series from

Table 2. Description of features.

Feature	F ₁	F ₂	F ₃	F ₄	F ₅	F ₆	F ₇	F ₈	F ₉	F ₁₀	F ₁₁	F ₁₂	F ₁₃	F ₁₄	F ₁₅
Description	X_p	\bar{X}	σ^2	S_{td}	X_{rms}	W_f	P_{kf}	I_f	S_{sk}	K_v	R_{f1}	R_{f2}	E_{svd}	H_{ISC1}	H_{ISC2}

**Figure 11.** LCD results of the signal under counter-position pistons abrasion.

nonlinear dynamic system.^{11,25,45} According to the definition of information entropy, the singular spectrum entropy using all of the ISCs from the original signal were extracted as features, which is given as

$$E_{svd} = - \sum_{i=1}^l p_i \ln p_i \quad (21)$$

where $p_i = \lambda_i / \sum_{i=1}^m \lambda_i$.

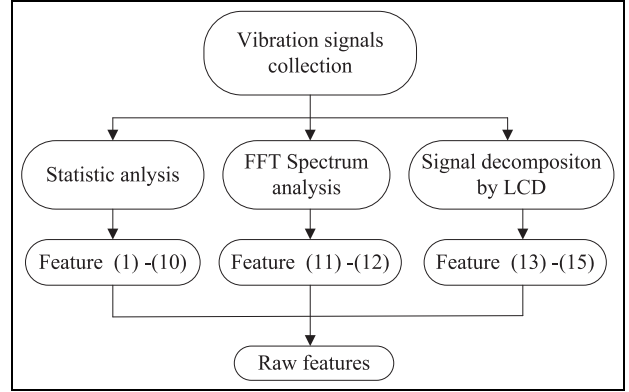
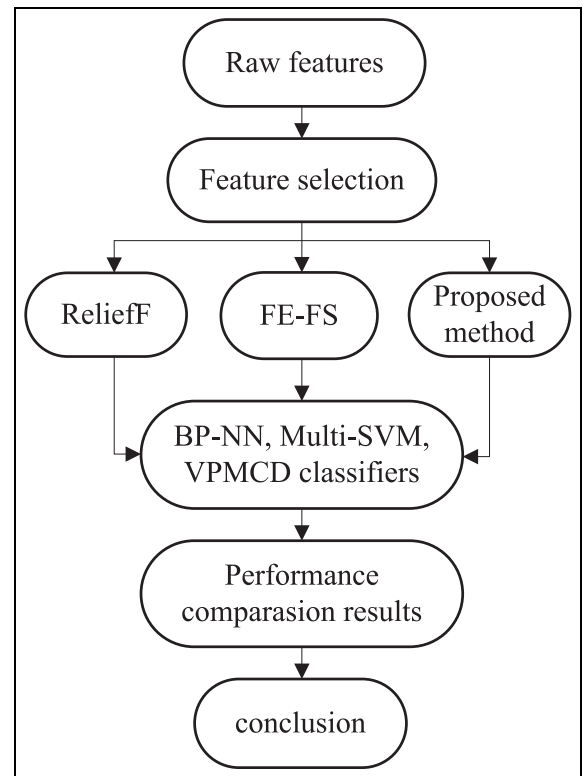
Simultaneously, it was found that the high-frequency ISCs were more informative. Therefore, the inherent fuzzy entropy values of the first two ISCs to quantify the fault-sensitive information as the 14th and 15th features:

14. The inherent fuzzy entropy value H_{ISC1} of the first ISC.
15. The inherent fuzzy entropy value H_{ISC2} of the second ISC.

In the end, 15 features were acquired. The flowchart of feature extraction is given in Figure 12. And, the description of features is shown in Table 2.

Fault diagnosis results and comparison

Three kinds of feature selection techniques and three kinds of classifiers were utilized to built fault diagnosis model and make a comparison. The flowchart of multi-fault diagnosis is shown in Figure 13. To begin with, 15 raw features were served as input vector and three different types of classification algorithms—Multi-SVM

**Figure 12.** Flowchart of feature extraction technique.**Figure 13.** Comparisons flowchart for multi-fault diagnosis of piston pump.

classifier, backpropagation neural network (BP-NN) classifier, and the improved predictive model-based class discrimination (PMCD) classifier—were employed, respectively. Different number of samples from each class were used as training samples to acquire the classifiers' model. The remaining samples

Table 3. Comparison results when different training samples and different classification techniques.

Features	Features dimension	Number of training samples	Classifier	Accuracy (%)			Time cost (s)
				Minimum	Mean	Standard deviation	
All	15	6	BP-NN	49.69	82.24	10.19	4.63
			Multi-SVM	95.06	99.97	0.31	6.82
			VPMCD	85.69	97.61	0.03	4.73
	10	10	BP-NN	83.00	86.50	4.95	4.87
			Multi-SVM	95.67	99.94	0.48	11.84
			VPMCD	96.67	99.32	0.01	5.10
	14	14	BP-NN	96.01	99.17	0.17	4.60
			Multi-SVM	100.00	100.00	0.00	21.83
			VPMCD	96.38	99.51	0.01	4.76
	20	20	BP-NN	45.42	99.80	1.89	6.66
			Multi-SVM	100	100	0	34.94
			VPMCD	96.67	99.71	0.01	4.59

BP-NN: back propagation neural network; SVM: support vector machine; VPMCD: variable predictive model-based class discrimination. The bold values show the shortcomings of BP-NN and Multi-SVM compared with VPMCD.

were employed to test the intelligent fault diagnosis model. Because the parameters of BP-NN can greatly affect the classification results, a BP-NN with three-layer structure of 15-8-6 was employed after analysis. The input of the network was raw vector which included 15 elements, and the output of the network was classification results responding to classes of the axial piston pump. The activation functions of input layer and output layer were “logsig” function and “line purelin” function, respectively. The training function was “traingdx” function. The training precise was set as 0.005 and the learning rate 0.001. As we all know, SVM is able to show consistent success rates, but SVM algorithm is binary (separating only two classes at a time) in nature, and the extension to multiclass problems requires iterative or combinatorial analysis using multiple classifiers and intensive optimization routines. Here, a Multi-SVM program was used with the linear kernel function pre-selected based on prior knowledge.

The test was done for 1000 times, and some results when using different training samples are presented in Table 3. When the number of training samples is less (such as 6), the mean accuracy of BP-NN model can get 82.24%, and it sounds promising, but minimum is 49.69% and the standard deviation is 10.19%, which is the highest in three methods. This shows that BP-NN model has poor robust. Although the mean and minimum accuracy of the VPMCD classifier seem weaker than the Multi-SVM model, VPMCD has higher stability. In addition, the VPMCD classifier costs much less time. When the number of training samples added up to 10, overall, all the three models do well. But BP-NN

model was not stable enough; it had the standard deviation at 4.95%, which was more than the one of the two other models. When the number of training samples was 14, the situation was roughly the same as above. When the number of training samples rises up to 20, the time consumption of Multi-SVM model was as long as 34.94 s. The minimum accuracy of BP-NN model cut down as low as 45.42%. On the contrary, the VPMCD classifier outperforms Multi-SVM at high accuracy, small standard deviation, and the lowest time consumption due to its clear mathematic expression and no iterative operation iteration. However, better performance of the VPMCD classifier is expected in engineering. Through analysis, it is found if the higher relevant features are selected to establish the mathematic model, the diagnosis performance can be improved.

As shown above that BP-NN classifier had weak stability, only multi-SVM and the improved VPMCD were used as class recognition methods to make a comparison between different feature selection techniques. First, the ReliefF approach served to accomplish the feature selection which ranked the features according to their importance weights,^{46,47} shown in Figure 14. From Figure 14, it can be seen that feature importance weights of number 2, 7, 8, and 14 features (four features in total below the dash line in Figure 14) were lower than 0.1 s; so, these features (below the dash line) were removed from the raw feature set. The results using Multi-SVM and VPMCD were illustrated in the first and the second rows in Table 4, respectively. Note that the results were attained when 10 training samples were used, and there was the same situation below.

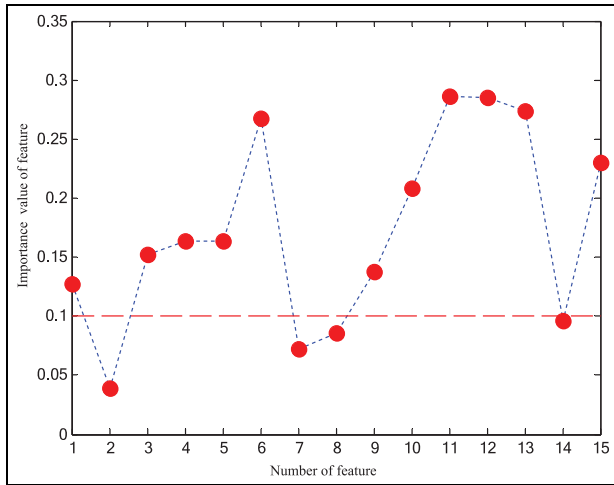


Figure 14. Importance value of features via ReliefF approach.

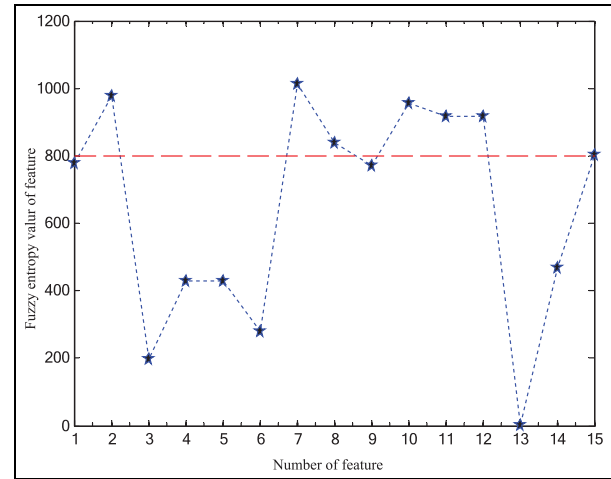


Figure 15. Fuzzy entropy values of features.

Second, the fuzzy entropy-based feature selection (FE-FS) approach^{22,35} was used to the same data sets to reduce the feature dimension, and the fuzzy entropy values calculated are visualized in Figure 15. Here, half selection strategy was taken to select approximately 50% of the features in the domain, so the number 2, 7, 8, 10, 11, 12, and 15 features (seven features above the dash line in Figure 15) were removed. The results using Multi-SVM and VPMCD were illustrated in the third and fourth rows in Table 4, respectively. The first four rows of Table 4 indicated that no matter which classification model was used, both the ReliefF and FE-FS can achieve better results and shortened time cost. Through further analysis for the first four rows, it can be shown that the improved VPMCD classifier outweighed Multi-SVM: faster classification speed and more robustness, which are vital for fault diagnosis in engineering practice.

Finally, the proposed feature selection method was utilized to select the optimal feature variables to diagnose the faults for the axial piston pump. Five optimal feature variables and the optimal $\{VPM_i^k, i = 1, 2, 3, 4, 5\} (k = 1, 2, \dots, 6)$ were obtained. In this stage, 360 samples were divided into two groups: 60 samples for

training and 300 samples for test. The procedure ended when the test mean accuracy decreased by more than 1% compared to the previous step or equaled to 100%, and the time consumption was less than 0.1 s. And, 300 test samples are served to verify the proposed fault diagnosis model. The predict errors of test samples are given in Figure 16. In Figure 16(a), while using the VPMs for class 1, the predict errors of samples Nos 1–50 are the lowest. This indicates that test samples Nos 1–50 belong to class 1, which is consistent with the actual experimental situation. The same situations exist in Figure 16(b)–(f), which is not repeated here.

Simultaneously, Multi-SVM was used to make a comparison. The results are given in the fifth and the sixth rows in Table 4. It can be shown that the Multi-SVM would take longer time due to intense computation though it can also achieve 100% accuracy and get excellent stability after using the proposed feature selection method. Whereas, the improved VPMCD can achieve the optimal diagnosis performance: the highest accuracy, the least time consumption (just 0.08 s), and the best robustness, which is vital for the condition monitoring and fault diagnosis online for key large-scale mechanical system.

Table 4. Comparison results of fault diagnosis of different strategies for the axial piston pump.

Feature selection technique	Selected features	Number of selected features	Classifier	Mean accuracy (%)	Standard deviation of accuracy (%)	Time cost (s)
ReliefF	[F ₁ , F ₃ , F ₄ , F ₅ , F ₆ , F ₉ , F ₁₀ , F ₁₁ , F ₁₂ , F ₁₃ , F ₁₅]	11	Multi-SVM	99.49	0.19	8.44
			Improved VPMCD	99.73	0.01	1.05
FE-FS	[F ₁ , F ₃ , F ₄ , F ₅ , F ₆ , F ₉ , F ₁₃ , F ₁₄]	8	Multi-SVM	99.72	0.94	4.84
			Improved VPMCD	99.35	0.014	0.73
Proposed method	[F ₅ , F ₁₀ , F ₁₁ , F ₁₃ , F ₁₄]	5	Multi-SVM	100	0	5.55
			Improved VPMCD	100	0	0.08

FE-FS: fuzzy entropy-based feature selection; SVM: support vector machine; VPMCD: variable predictive model-based class discrimination. The bold values show the superiority of VPMCD.

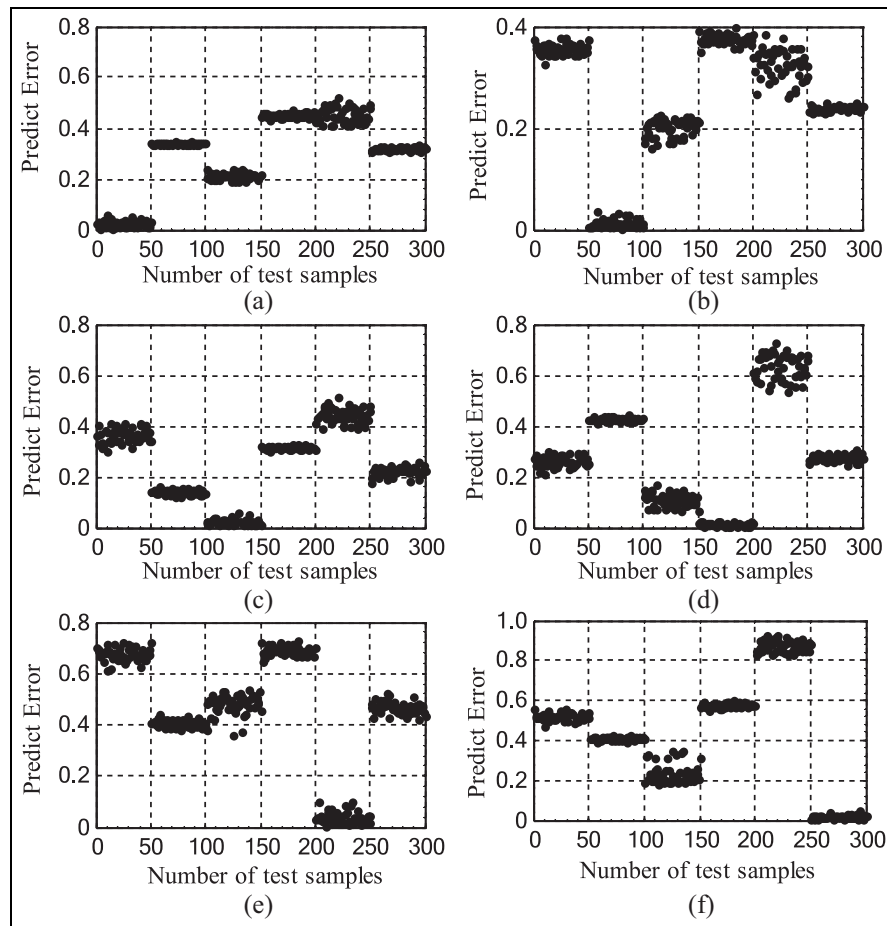


Figure 16. Predict errors of test samples using the VPMs: (a) for class 1, (b) for class 2, (c) for class 3, (d) class 4, (e) for class 5, and (f) for class 6.

Conclusion

First, an improve VPMCD algorithm was proposed to adapt the non-normal distribution characteristics of features. Besides, a novel feature selection technique has been developed by integrating the similarity-fuzzy entropy and the improved VPMCD method. The proposed feature selection technique can find the ideal feature vector using FCM approach for the similarity measures estimation and obtain more accurate fuzzy entropy values, so the dimension of input feature variables can be effectively reduced to boost the performance during the VPM training and test procedure of the VPMCD classifiers. At last, the proposed method was applied to establish incipient intelligent multi-fault diagnosis model for the axial piston pump, which includes three types of single-fault and two types of multi-fault. The results demonstrate that the proposed technique can sort out the most relevant features as optimal features which are more sensitive to fault type and more adaptive. The improved VPMCD classifiers with the optimal features have achieved higher accuracy, less computational cost, and best robustness. But, it should be noted that the VPMCD classifiers need to pre-determine the type of VPMs, which likely leads to unstability of classification results when different

models are used. Although this weakness has been considered in this paper, further research should be done in the future.

Acknowledgement

The authors greatly appreciate the support from Cooperative Innovation Center for the Construction and Development of Dongting Lake Ecological Economic Zone, and China Scholarship Council.

Data availability

All data included in this study are available upon request by contacting the corresponding author.

Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Natural Science Foundation of Hunan Province (2018JJ2275, 2019JJ6002),

Scientific Research Foundation of Hunan Provincial Education Department (17A147), Doctoral Scientific Research Start-up Foundation of Hunan University of Arts and Science (16BSQD22), and National Natural Science Foundation of China (11402036).

ORCID iD

Songrong Luo  <https://orcid.org/0000-0002-8440-2300>

References

- Hoang D-T and Kang HJ. Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cogn Syst Res* 2019; 53: 42–50.
- Jia F, Lei Y, Guo L, et al. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing* 2018; 272: 619–628.
- Yang Y, Wang H, Cheng J, et al. A fault diagnosis approach for roller bearing based on VPMCD under variable speed condition. *Measurement* 2013; 46(8): 2306–2312.
- Liu R, Yang B, Zio E, et al. Artificial intelligence for fault diagnosis of rotating machinery: a review. *Mech Syst Signal Proc* 2018; 108: 33–47.
- Liu X and Bo L. Identification of resonance states of rotor-bearing system using RQA and optimal binary tree SVM. *Neurocomputing* 2015; 152: 36–44.
- Jia F, Lei Y, Lu N, et al. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mech Syst Signal Proc* 2018; 110: 349–367.
- Jiang X, Wang J, Shi J, et al. A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines. *Mech Syst Signal Proc* 2019; 116: 668–692.
- Sahani M and Dash PK. Variational mode decomposition and weighted online sequential extreme learning machine for power quality event patterns recognition. *Neurocomputing* 2018; 310: 10–27.
- Mahela OP and Shaik AG. Recognition of power quality disturbances using S-transform based ruled decision tree and fuzzy C-means clustering classifiers. *Appl Soft Comput* 2017; 59: 243–257.
- Li Z, Yan X, Wang X, et al. Detection of gear cracks in a complex gearbox of wind turbines using supervised bounded component analysis of vibration signals collected from multi-channel sensors. *J Sound Vib* 2016; 371(9): 406–433.
- Huang J, Wang X, Wang D, et al. Analysis of weak fault in hydraulic system based on multi-scale permutation entropy of fault-sensitive intrinsic mode function and deep belief network. *Entropy* 2019; 21(4): 425.
- Berredjem T and Benidir M. Bearing faults diagnosis using fuzzy expert system relying on an improved range overlaps and similarity method. *Expert Syst Appl* 2018; 108: 134–142.
- Hemmati F, Orfali W and Gadala MS. Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation. *Appl Acoust* 2016; 104: 101–118.
- Lei Y, Lin J, He Z, et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech Syst Signal Proc* 2013; 35: 108–126.
- Zhang X and Zhou J. Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. *Mech Syst Signal Proc* 2013; 41(1–2): 127–140.
- Lan Y, Hu J, Huang J, et al. Fault diagnosis on slipper abrasion of axial piston pump based on extreme learning machine. *Measurement* 2018; 124: 378–385.
- Ali JB, Fnaiech N, Saidi L, et al. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl Acoust* 2015; 89: 16–27.
- Saravanan N and Ramachandran KI. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). *Expert Syst Appl* 2010; 89: 4168–4181.
- Tang T, Bo L, Liu X, et al. Variable predictive model class discrimination using novel predictive models and a adaptive feature selection for bearing fault identification. *J Sound Vib* 2018; 425: 137–148.
- Sheikhpour R, Sarram MA, Gharaghani S, et al. A survey on semi-supervised feature selection methods. *Pattern Recogn* 2017; 64: 141–158.
- Jaganathan P and Kuppuchamy R. A threshold fuzzy entropy based feature selection for medical database classification. *Comput Biol Med* 2013; 43: 2222–2229.
- Lohrmann C, Luukka P, Jablonska-Sabuka M, et al. A combination of fuzzy similarity measures and fuzzy entropy measures for supervised feature selection. *Expert Syst Appl* 2018; 110: 216–326.
- Chandrashekar G and Sahin F. A survey on feature selection methods. *Comput Electr Eng* 2014; 40: 16–28.
- Tang B, Song T, Li F, et al. Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. *Renew Energ* 2014; 62: 1–9.
- Li Y, Yang Y, Wang X, et al. Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine. *J Sound Vib* 2018; 428(218): 72–86.
- Raghuraj R and Lakshminarayanan S. Variable predictive models—a new multivariate classification approach for pattern recognition applications. *Pattern Recogn* 2009; 42(1): 7–16.
- Raghuraj R and Lakshminarayanan S. Variable predictive model based classification algorithm for effective separation of protein structural classes. *Comput Biol Chem* 2008; 32(4): 302–306.
- Raghuraj R and Lakshminarayanan S. VPMCD: variable interaction modeling approach for class discrimination in biological systems. *FEBS Lett* 2007; 581(5–6): 826–830.
- Luo S, Cheng J and Yang Y. An intelligent fault diagnosis method for rotating machinery based on multi-scale higher order singular spectrum analysis and GA-VPMCD. *Measurement* 2016; 87: 38–50.
- Yang Y, Pan H, Ma L, et al. A fault diagnosis approach for roller bearing based on improved intrinsic timescale decomposition de-noising and kriging-variable predictive model-based class discriminate. *J Vib Control* 2016; 22(5): 1431–1446.

31. Zheng J, Cheng J, Yang Y, et al. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mech Mach Theory* 2014; 78: 187–200.
32. Luo S, Cheng J and Wei K. A fault diagnosis model based on LCD-SVD-ANN-MIV and VPMCD for rotating machinery. *Shock Vib* 2016; 2016: 5141564.
33. Cui H, Hong M, Qiao Y, et al. Application of VPMCD method based on PLS for rolling bearing fault diagnosis. *J Vibroeng* 2017; 19(1): 160–174.
34. Miao J and Niu L. A survey on feature selection. *Procedia Comput Sci* 2016; 91: 919–926.
35. Luukka P. Feature selection using fuzzy entropy measures with similarity classifier. *Expert Syst Appl* 2011; 38: 4600–4607.
36. Parkash O, Sharma PK and Mahajan R. New measures of weighted fuzzy entropy and their applications for the study of maximum weighted fuzzy entropy principle. *Inform Sciences* 2008; 178: 2389–2395.
37. Ma Z, Wang S, Shi J, et al. Fault diagnosis of an intelligent hydraulic pump based on a nonlinear unknown input observer. *Chinese J Aeronaut* 2018; 31(2): 385–394.
38. Li H, Sun J, Ma H, et al. A novel method based upon modified composite spectrum and relative entropy for degradation feature extraction of hydraulic pump. *Mech Syst Signal Pr* 2019; 114: 399–412.
39. Du J, Wang S and Zhang H. Layered clustering multi-fault diagnosis for hydraulic piston pump. *Mech Syst Signal Pr* 2013; 36: 487–504.
40. Azadeh A, Saberi M, Kazem A, et al. A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. *Appl Soft Comput* 2013; 13(3): 1478–1485.
41. Tian H, Wu Y, Qiang P, et al. An automatic abrupt information extraction method based on singular value decomposition and higher-order statistics. *Meas Sci Technol* 2016; 27(2): 025007.
42. Zheng J, Cheng J and Yang Y. A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy. *Mech Mach Theory* 2013; 70: 441–453.
43. Yang W. Establishment of the mathematical model for diagnosing the engine valve faults by genetic programming. *J Sound Vib* 2006; 293(1–2): 213–226.
44. Cheng J, Yu D, Tang J, et al. Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery. *Shock Vib* 2009; 16: 89–98.
45. Hortelano M, Reilly RB, Castells F, et al. Refined multi-scale fuzzy entropy to analyse post-exercise cardiovascular response in older adults with orthostatic intolerance. *Entropy* 2018; 20: 1–12.
46. Robnik-Sikonja M and Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn J* 2003; 53: 23–69.
47. Beretta L and Santaniello A. Implementing ReliefF filters to extract meaningful features from genetic lifetime datasets. *J Biomed Inform* 2011; 44: 361–369.